

Analisis Data *E-Absensi* untuk Menganalisis Perbandingan Pola Disiplin Kerja menggunakan Algoritma *Clustering K-Means*

¹Reksa Suhud Tri Atmojo, ²Leady Pramudita, ³Heni Novita Sari, ⁴Bagus Arifudin, ⁵Gigih Forda Nama

¹Program Studi Teknik Informatika
Jurusan Teknik Elektro, Fakultas Teknik, Universitas Lampung
Jalan Soemantri Brojonegoro No.1, Bandar Lampung, 35141
1reksa.suhudtri1501@students.unila.ac.id

Intisari — Data mining merupakan suatu proses untuk menemukan informasi-informasi yang bermanfaat dari sekumpulan basis data yang berukuran besar. Pada penelitian ini akan mengolah data yang ada pada sistem e-absensi menggunakan algoritma K-Means K = 3 menggunakan perangkat lunak Orange. Hasil penelitian ini menghasilkan hasil analisis pola perilaku kedisiplinan kerja dengan melihat presentase telat dan jumlah kehadiran yang diambil dari data E-Absensi SD Negeri X Taman Fajar Kecamatan Purbolinggo dan Puskesmas Y yang berada di Desa Negara Nabung, Kabupaten Lampung Timur sehingga dapat menjadi pertimbangan bagi instansi terkait peningkatan tingkat kedisiplinan kerja. Dari penelitian, ini didapatkan kesimpulan pada clustering C1, C2, dan C3 dapat disimpulkan tingkat kedisiplinan kerja pada saat Bulan Ramadhan 1438 H yang jatuh pada bulan Juli 2018 di pegawai Puskesmas Y lebih tinggi dibandingkan dengan pegawai SDN X Taman Fajar dari segi jumlah kehadiran, namun jika dilihat dari tingkat kedisiplinan dalam aspek presentase telat, Pegawai SDN X Taman Fajar lebih rendah dibanding pegawai Puskesmas Y.

Kata Kunci — Clustering, K-Means, Algoritma, E-Absensi, Pegawai

Abstract — Data mining is a process for finding useful information from a large collection of databases. In this study, we will process the data in the e-attendance system using the K-Means K = 3 algorithm using Orange software. The results of this study resulted in the analysis of patterns of work discipline behavior by looking at the late percentage and number of attendance taken from the data of Time Attendance at SD Negeri X Taman Fajar in Purbolinggo Subdistrict and Y Public Health Center in Negara Nabung Village, East Lampung Regency so that it could be considered by agencies related to an increase in the level of employment discipline. From this research, it can be concluded that clustering C1, C2, and C3 can be concluded that the level of work discipline at the time of Ramadhan 1438 H which falls in July 2018 at Y Public Health Center employees is higher than SDN X Taman Fajar employees in terms of attendance, but when viewed from the level of discipline in the late percentage aspect, SDN X Taman Fajar employees are lower than Y Puskesmas employees.

Keywords — Clustering, K-Means, Algorithm, E-Absensi, Employee

I. PENDAHULUAN

Setiap organisasi menginginkan pegawainya dapat bekerja secara maksimal mempengaruhi kinerja seorang pegawai, seperti umur, jenis kelamin, absensi. Absensi itu sendiri ialah tingkat kehadiran pegawai yang berkenaan dengan tanggung jawab. Kehadiran berkenaan

dengan tanggung jawab pegawai saat bekerja, pegawai yang hadir tepat waktu, tidak terlambat dan tidak sering izin saat masuk kerja bisa dikatakan mempunyai sifat disiplin. Dalam rangka meningkatkan penertiban administrasi kepegawaian khususnya pencatatan kehadiran pegawai dinas di Lampung Timur, setiap Pegawai Negeri Sipil harus masuk kerja dan menaati jam kerja.

Metode K-Means digunakan untuk mengelompokan absensi pegawai per bulan. Banyaknya jumlah transaksi absensi Pegawai Negeri Sipil yang dihasilkan dengan keterangan absensi masuk saja, absensi pulang saja, sering cuti dan sakit dapat dikelompokan menggunakan metode K-Means dengan menggunakan perangkat lunak pengolahan data mining yang bernama Orange. Hasil penelitian ini menghasilkan hasil analisis pola perilaku kedisiplinan kerja dengan melihat presentase telat dan jumlah kehadiran yang diambil dari data E-Absensi SD Negeri X Taman Fajar Kec. Purbolinggo dan Puskesmas Y yang berada di Desa Negara Nabung, Kabupaten Lampung Timur.

II. TINJAUAN PUSTAKA

A. Pengertian Data Mining

Tan (2006) mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery*. Salah satu teknik yang dibuat dalam *data mining* adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Dalam *data mining*, pengelompokan data juga bisa dilakukan. Tujuannya adalah agar kita dapat mengetahui pola universal data-data yang ada. Anomali data transaksi juga perlu dideteksi untuk dapat mengetahui tindak lanjut berikutnya yang dapat diambil [1].

B. Pengertian K-Means

Merupakan mengelompokan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokan

dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil. Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut:

- Pilih jumlah *cluster* k .
- Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara *random*. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka *random*,
- Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{in} - X_{jn})^2} \dots (1)$$

dimana:

$D(i, j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

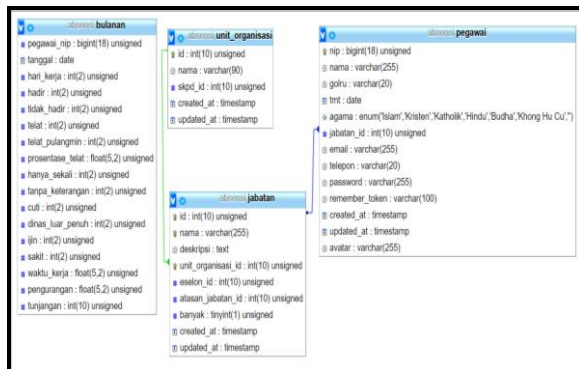
(1)

- Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
- Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat cluster tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi [2].

III. METODOLOGI PENELITIAN

A. Melakukan Preprocessing Data

Dikarenakan data yang diperoleh dari E-Absensi Lampung Timur dalam bentuk database SQL maka perlu dilakukan normalisasi join table agar data yang diperoleh mudah untuk diolah. Berikut adalah struktur *database* aplikasi E-Absensi yang dianalisis



Gbr 1 . Struktur relasi database sistem e-absensi kepegawaian Kabupaten Lampung Timur

Selanjutnya, melakukan *filtering data* menggunakan query SQL karena data SD Negeri X Taman Fajar Kec. Purbolingo dan Puskesmas Y yang berada di Desa Negara Nabung, Kabupaten Lampung Timur yang ingin dianalisa.

```

SELECT pegawai.`nama` AS
nama_pegawai, unit_organisasi.`nama` AS
nama_instansi,
bulanan.`prosentase_telat` AS
presentase_telat, bulanan.`hadir`,
bulanan.`tidak_hadir`
FROM pegawai JOIN bulanan ON
bulanan.`pegawai_nip`=pegawai.`nip`
JOIN jabatan ON
jabatan.`id`=pegawai.`jabatan_id`
JOIN unit_organisasi ON
unit_organisasi.`id`=jabatan.`id`
  
```

Berikut ini adalah hasil dari join tabel dengan query SQL :

Atribut yang akan dianalisis adalah nama, organisasi, presentase telat, hadir, dan tidak hadir. Dikarenakan yang dianalisis adalah kehadiran maka yang ditandai label adalah atribut hadir. Total record data salah satu

SDN dan Puskesmas di Kabupaten Lampung Timur yang akan diproses menggunakan K-Means Clustering pada penelitian ini ada 55 Record.

Tabel 1 Tabel hasil join

nama	organisasi	presentase telat	hadir	tidak hadir
R	P	7,37	21	0
A	P	7,38	21	0
H	P	7,4	21	0
A	P	7,37	19	2
S	P	7,63	21	0
I	P	7,69	21	0
A	P	7,71	20	1
	P	2,14	15	6
	P	2,71	18	3
	P	2,89	20	1
	P	2,89	20	1
	P	2,89	20	1
	P	3,02	20	1
	P	3,05	20	1
	P	0	0	21
	P	0	0	21
	P	3,04	20	1
	P	3,05	20	1
	P	2,31	18	3
	P	0	0	21
	P	3,07	20	1
	P	2,92	20	1
	P	3,98	16	5
	P	0	0	21
	P	3,09	20	1
	P	3,1	20	1
	P	2,59	11	10
	P	0	0	21
	P	3,01	20	1
	P	1,91	11	10
	P	4,26	20	1
	P	3,06	20	1
	P	3,06	20	1
	P	3,24	19	2
	P	3,26	20	1
	P	7	15	6
	P	1,23	1	20
	P	0	0	21
	P	1,21	1	20
	P	3,25	20	1
	P	3,25	20	1
	P	3,25	20	1
	P	3,25	20	1
	P	0	0	21
	P	0,1	3	18
	P	3,11	20	1
	P	2,42	12	9
	P	3,14	20	1
	P	3,14	20	1
	P	3,11	20	1
	P	3,15	20	1
	P	0	0	21
	P	2,77	11	10
	P	0	0	21

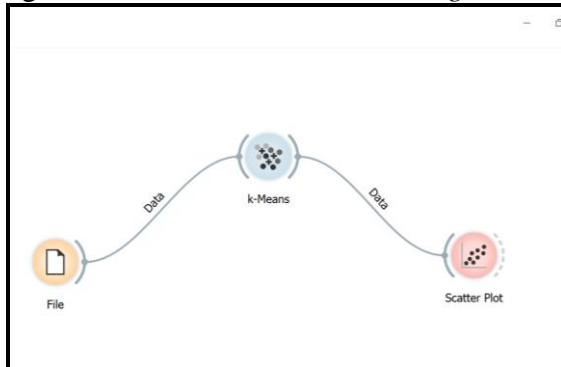
B. Melakukan clustering K-Means

Clustering K-Means bertujuan untuk melakukan *clustering* pada pegawai SD Negeri XTaman Fajar Kec. Purbolingo dan Puskesmas Y yang berada di Desa Negara Nabung, Kabupaten Lampung Timur yang ingin dianalisa. *Clustering* yang kami gunakan adalah K = 3.

C. Melakukan output data menggunakan penyajian data scatter plot

Karena data yang diolah menggunakan *algoritma K-Means* dengan *Clustering K = 3* dengan *maximum iteration* sebanyak 300 iterasi pada perangkat lunak *orange*. Maka

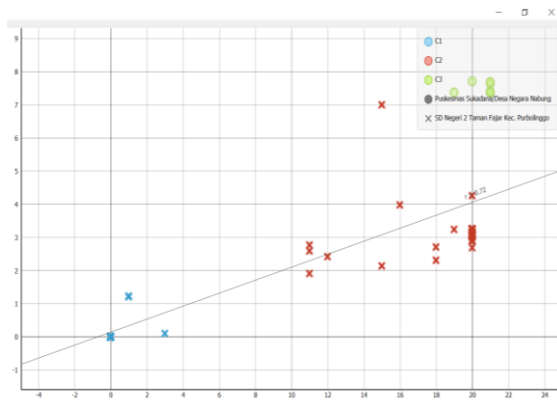
akan menghasilkan data C1, C2, dan C3 yang akan disajikan menggunakan *scatter plot*. Secara garis besar berikut proses clustering yang dilakukan melalui *canvas Orange*



Gb 2. Struktur proses canvas clustering k-means pada perangkat lunak Orange

IV. PEMBAHASAN

Setelah metode penelitian sudah dilakukan dan diolah melalui perangkat lunak *Orange*, didapatkan data yang sudah di sajikan melalui *Scatter Plot*.



Gbr 3 . Hasil clustering K-Means yang disajikan dalam bentuk scatter plot

Pada Gambar 3 Sumbu X adalah data hadir dan Sumbu Y adalah data presentase telat hadir. Dari hasil pengolahan menggunakan Clustering didapatkan C1, C2, dan C3. C1 merupakan data pegawai yang memiliki intensitas kehadiran rendah dan presentase telat rendah. Clustering C2 diasumsikan sebagai pegawai yang memiliki intensitas kehadiran dan presentase telat rata- yang diasumsikan sebagai pegawai yang memiliki kehadiran rendah dan presentase telat rendah. Sedangkan clustering C3 diasumsikan sebagai pegawai yang memiliki intensitas kehadiran tinggi dan presentase telat yang

tinggi. *Clustering C1* mayoritas datanya diisi oleh data pegawai SDN X Taman Fajar yang memiliki kehadiran rendah namun dengan presentase telat rendah pada saat bulan Ramadhan 1438H dengan centroid cluster $K_1 = 0,72$. Dapat disimpulkan hubungan adalah berbanding lurus, semakin rendah jumlah kehadiran maka semakin rendah presentase telat. *Clustering C2* mayoritas datanya diisi oleh data pegawai SDN X Taman Fajar yang memiliki kehadiran rata-rata dan dengan presentase telat rata-rata pada absensi saat bulan Ramadhan 1438H dan centroid cluster $K_2 = 0,79$. Dapat disimpulkan hubungan adalah berbanding lurus, semakin rendah jumlah kehadiran maka semakin rendah presentase telat. *Clustering C3* mayoritas datanya diisi oleh data pegawai Puskesmas Y yang memiliki kehadiran tinggi namun dengan presentase telat yang tinggi pula pada saat bulan Ramadhan 1438H dengan centroid cluster $K_3 = 0,767$. Dapat disimpulkan hubungan adalah berbanding lurus, semakin tinggi jumlah kehadiran maka semakin tinggi presentase telat.

Dari ketiga buah *clustering C1, C2, dan C3* dapat disimpulkan tingkat kedisiplinan kerja pada saat Bulan Ramadhan 1438H yang jatuh pada bulan juli 2018 pada pegawai Puskesmas Y lebih tinggi dibandingkan dengan pegawai SDN X Taman Fajar dari segi jumlah kehadiran, namun jika dilihat dari tingkat kedisiplinan dalam aspek presentase telat, Pegawai SDN X Taman Fajar lebih rendah dibanding pegawai Puskesmas Sukadana.

Hal ini dapat menjadi pertimbangan pihak SDN X Taman Fajar mengapa tingkat kedisiplinan dilingkungan kerja SDN X Taman Fajar rendah, ataupun bisa jadi faktor lain, misalnya ada yang menyebabkan data kehadiran tidak tercatat di e-absensi dikarenakan kurangnya penyuluhan cara penggunaan system absensi.

V. KESIMPULAN

Dari hasil Penelitian didapatkan kesimpulan sebagai berikut :

- Metode data mining pada system e-absensi pada SDN X Taman Fajar dan Puskesmas Y menggunakan algoritma

Clustering K-Means $K = 3$ dan menghasilkan cluster C1, C2, dan C3.

- Hubungan antara jumlah kehadiran dan presentase telat adalah berbanding lurus pada clustering C1, C2, dan C3. Dimana semakin tinggi jumlah kehadiran maka presentase telat juga berpotensi semakin tinggi, begitu juga sebaliknya
- Dari ketiga buah clustering C1, C2, dan C3 dapat disimpulkan tingkat kedisiplinan kerja pada saat Bulan Ramadhan 1438 H pada pegawai Puskesmas Y lebih tinggi dibandingkan dengan pegawai SDN X Taman Fajar dari segi jumlah kehadiran, namun jika dilihat dari tingkat kedisiplinan dalam aspek presentase telat, Pegawai SDN X Taman Fajar lebih

rendah dibanding pegawai Puskesmas Sukadana.

REFERENSI

- [1] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama Vol.*, vol. 11, no. 2, pp. 130–138, 2015.
- [2] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.